3-14-2019

# Four Draft Single-Cell Genome Sequences of Novel, Nearly Identical Kiritimatiellaeota Strains Isolated from the Continental Deep Subsurface

Joshua D. Sackett
*Desert Research Institute Las Vegas*

Brittany R. Kruger
*Desert Research Institute Las Vegas*

Eric D. Becraft
*Bigelow Laboratory for Ocean Sciences*

Jessica K. Jarett
*U.S. Department of Energy Joint Genome Institute*

Ramunas Stepanauskas
*Bigelow Laboratory for Ocean Sciences*

*See next page for additional authors*



## Recommended Citation

Citation Sackett JD, Kruger BR, Becraft ED, Jarett JK, Stepanauskas R, Woyke T, Moser DP. 2019. Four draft single-cell genome sequences of novel, nearly identical Kiritimatiellaeota strains isolated from the continental deep subsurface. Microbiol Resour Announc 8:e01249-18. https://doi.org/10.1128/MRA.01249-18.

AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

# Four Draft Single-Cell Genome Sequences of Novel, Nearly Identical *Kiritimatiellaeota* Strains Isolated from the Continental Deep Subsurface

Joshua D. Sackett,[a,b,c] Brittany R. Kruger,[a,b] Eric D. Becraft,[d,e] Jessica K. Jarett,[f] Ramunas Stepanauskas,[d] Tanja Woyke,[f] Duane P. Moser[a,b]

[a]Division of Earth and Ecosystems Sciences, Desert Research Institute, Las Vegas, Nevada, USA
[b]Division of Hydrologic Sciences, Desert Research Institute, Las Vegas, Nevada, USA
[c]School of Life Sciences, University of Nevada Las Vegas, Las Vegas, Nevada, USA
[d]Bigelow Laboratory for Ocean Sciences, East Boothbay, Maine, USA
[e]Department of Biology, University of North Alabama, Florence, Alabama, USA
[f]Joint Genome Institute, Walnut Creek, California, USA

**ABSTRACT** The recently proposed bacterial phylum *Kiritimatiellaeota* represents a globally distributed monophyletic clade distinct from other members of the *Planctomycetes*, *Verrucomicrobia*, and *Chlamydiae* (PVC) superphylum. Here, we present four phylogenetically distinct single-cell genome sequences from within the *Kiritimatiellaeota* lineage sampled from deep continental subsurface aquifer fluids of the Death Valley Regional Flow System in the United States.

Members of the recently proposed bacterial phylum *Kiritimatiellaeota* (1) (previously *Verrucomicrobia* subdivision 5 [2]) are globally distributed and found in environments such as vertebrate intestines (3), soils (4), and marine environments (1, 5, 6). However, despite their cosmopolitan distribution and prevalence in 16S rRNA gene amplicon surveys, little is known about the genomic diversity, physiology, and ecology of these organisms, particularly in deep continental subsurface environments.

To date, a single pure culture representative of the *Kiritimatiellaeota* (*Kiritimatiella glycovorans* L21-Fru-AB$^T$), originally isolated from a hypersaline lake on the Kiritimati Atoll, has been cultivated and phenotypically and genomically characterized (1, 6). In line with previous observations of polysaccharide degradation by members of this group (5), cultivation studies and genomic analysis of *K. glycovorans* L21-Fru-AB$^T$ suggest that this organism is saccharolytic and derives energy via fermentation (1). Here, we report four draft single-cell genome sequences representing members of the *Kiritimatiellaeota* phylum obtained from a deep, fractured rock aquifer.

Subsurface aquifer water samples were collected with a motor-driven discrete sampler from an uncased interval at a depth of 752 m below the land surface in BLM1, an 883.5-m-deep monitoring borehole drilled into Paleozoic carbonates located in Inyo County, California (36.4004°N, −116.4692°W), in August 2015. The water temperature was 57.2°C, the pH was 6.92, the electrical conductivity was 2,299 $\mu$S cm$^{-1}$, and the oxidation-reduction potential was −242 mV. Despite a dissolved oxygen measurement of 0.43 mg liter$^{-1}$, the downhole environment was most likely anoxic owing to its negative oxidation-reduction potential. Raw water samples (1 ml) for single-cell genomics were amended with 5% glycerol and 1× Tris-EDTA (TE) buffer (final concentrations), frozen on dry ice in the field, and stored at −80°C until cell sorting. Single cells were sorted, and their genomes were amplified and sequenced at the Bigelow Laboratory for Ocean Sciences Single Cell Genomics Center as previously described (7). Briefly, cryopreserved samples were thawed, prescreened through a 40-$\mu$m nylon mesh cell

**Citation** Sackett JD, Kruger BR, Becraft ED, Jarett JK, Stepanauskas R, Woyke T, Moser DP. 2019. Four draft single-cell genome sequences of novel, nearly identical *Kiritimatiellaeota* strains isolated from the continental deep subsurface. Microbiol Resour Announc 8:e01249-18. https://doi.org/10.1128/MRA.01249-18.

**Editor** Vincent Bruno, University of Maryland School of Medicine



Address correspondence to Duane P. Moser, duane.moser@dri.edu.

**Received** 21 September 2018
**Accepted** 30 January 2019
**Published** 14 March 2019

strainer (Becton Dickinson, Franklin Lakes, NJ, USA), and incubated with SYTO-9 DNA stain (Thermo Fisher Scientific, Waltham, MA, USA) at a final concentration of 5 $\mu$M for 10 to 60 min. Fluorescence-activated cell sorting was performed with a BD InFlux Mariner flow cytometer equipped with a 488-nm laser and a 70-$\mu$m nozzle orifice (Becton Dickinson). The cytometer was triggered on side scatter, and the "single-1 drop" mode was used for maximal sort purity. The sort gate was defined based on particle green fluorescence, light side scatter, and the ratio of green versus red fluorescence (for improved discrimination of cells from detrital particles). For each sample, individual cells were deposited into 384-well plates containing 600 nl per well of 1× TE buffer and stored at −80°C prior to subsequent processing. Of the 384 wells, 317 wells were dedicated for single particles, 64 wells were used as negative controls (no droplet deposition), and 3 wells received 10 particles each to serve as positive controls. Cells were lysed, and their DNA was denatured with 5 freeze-thaw cycles, the addition of 700 nl of lysis buffer (0.4 M KOH, 10 mM EDTA, and 100 mM dithiothreitol), and a subsequent 10-min incubation at 20°C. Lysis was terminated by the addition of 700 nl of 1 M Tris-HCl at pH 4.

Sequencing libraries were created for each single cell with the Nextera XT DNA library preparation kit (Illumina, San Diego, CA, USA) with the following modifications: purification was performed with column cleanup kits (Qiagen, Venlo, the Netherlands), and library selection was performed with BluePippin (Sage Science, Beverly, MA, USA) with a target sequence size of 500 ± 50 bp. Libraries were sequenced with the NextSeq 500 platform (Illumina) and V1 reagents (2 × 150-bp paired-end sequencing). Raw sequencing reads for each single amplified genome (SAG) were quality trimmed with Trimmomatic v0.32 (8), reads with 95% or greater nucleotide identity with the *Homo sapiens* reference genome assembly (GRCh38) were removed, and low-complexity reads (less than 5% of any nucleotide) were removed as described previously (7). Quality-filtered reads were normalized *in silico* with kmernorm 1.05 (http://sourceforge.net/projects/kmernorm) using the settings –k 21 –t 30 –c 3 and subsequently assembled into contigs with SPAdes v3.9.0 (9) with the following settings: –careful –sc –phred-offset 33. Contig ends (100 bp) were trimmed, and contigs of fewer than 2,000 bp were discarded. Genome completeness and potential contamination were estimated with CheckM v1.0.8 (10). Predicted genome size was calculated by dividing assembly size by estimated genome completeness. Assembly quality for each SAG was determined according to minimum information about single amplified genome (MISAG) standards (11). Protein-encoding regions were identified with the Rapid Annotations using Subsystems Technology (RAST) server (12), and genes were annotated with Koala (KEGG) (13) and InterProScan 5 (14). Average nucleotide identity (ANI) of reciprocal hits between genome assemblies was calculated using the online ANI calculator (http://enve-omics.ce.gatech.edu/ani/) (15). Assembly statistics are shown in Table 1.

Based on the detection of conserved single-copy marker genes in the 3 most complete SAG assemblies, we predict that BLM1 *Kiritimatiellaeota* genome sequences contain 3.8 to 4.2 Mbp. The CheckM-based predicted genome size of the smallest SAG (AH-151-K23) was 3 times higher than values for the other SAGs. CheckM estimates genome completeness and contamination of genome assemblies based on the presence and location of lineage-specific marker genes selected from the phylogenetic placement (based on single-copy marker genes in the assembly) of the assembly in a built-in reference genome tree (10). Of the 104 marker genes used by CheckM to assess genome completeness and contamination for AH-151-K23, only 2 genes were found in the assembly (threonylcarbamoyl adenosine biosynthesis protein TsaE [accession no. PF02367] and Holliday junction DNA helicase RuvA [accession no. TIGR00084]), ultimately resulting in 3.4% estimated genome completeness. Furthermore, compared to the other SAGs, the largest contig size (30 kb) and $N_{50}$ value (7.7 kb) associated with AH-151-K23 were ~2 to 5 times lower. A combination of the absence of phylogenetically informative marker genes in the assembly, low genome recovery (small assembly),

**TABLE 1** Assembly and quality statistics for BLM1 *Kiritimatiellaeota* SAGs

| Assembly statistic | Data for SAG: AH-151-K23 | AH-147-K21 | AH-151-C14 | AH-151-A22 |
|---|---|---|---|---|
| Raw read accession no. | SAMEA5244953 | SAMEA5244950 | SAMEA5244951 | SAMEA5244952 |
| Assembly accession no. | CAACVW010000000 | CAACVX010000000 | CAACVZ010000000 | CAACVY010000000 |
| Annotation accession no. | 3300022259 | 3300022272 | 3300022292 | 3300022301 |
| No. of raw paired-end reads | 9,035,693 | 4,110,878 | 8,955,047 | 7,063,659 |
| No. of quality-filtered paired-end reads | 222,128 | 312,138 | 731,454 | 1,107,044 |
| Assembly size (bp) | 431,650 | 688,295 | 1,629,841 | 2,632,675 |
| G+C content (%) | 61.5 | 61.6 | 61.8 | 62.1 |
| Estimated genome completeness (%)[a] | 3.4 | 17.4 | 38.9 | 69.0 |
| Predicted genome size (Mbp)[a] | 12.52 | 3.95 | 4.19 | 3.82 |
| Estimated contamination (%)[a] | 0 | 0.7 | 0.9 | 0 |
| Genome quality[b] | Low | Low | Low | Medium |
| No. of contigs | 73 | 54 | 161 | 137 |
| Largest contig (bp) | 30,074 | 56,953 | 91,734 | 150,914 |
| $N_{50}$ value | 7,761 | 20,490 | 16,267 | 37,970 |
| No. of protein-coding genes | 409 | 595 | 1,394 | 2,191 |
| No. of tRNA genes | 5 | 15 | 25 | 39 |
| No. of rRNA genes | 4 | 3 | 3 | 3 |

[a] Estimated with CheckM v1.0.8 (10).
[b] Genome quality reported according to Bowers et al. (11).

and relatively short contigs contributed to very low genome completeness and high genome size predictions for this SAG.

All four SAGs had identical 16S rRNA genes and shared greater than 99% average nucleotide identity. The 16S rRNA gene sequence has 82.7% sequence identity with *K. glycovorans* L21-Fru-AB$^T$ (GenBank accession no. KC665948) (1), suggesting that these 2 organisms belong to genetically distinct lineages. The SAGs encode a variety of glycosyl hydrolases, including cellulases (GH5), $\beta$-xylosidases (GH39), D-4,5-unsaturated $\beta$-glucuronyl hydrolases (GH88), glucoamylases (GH97), and endo-$\alpha$-*N*-acetylgalactosaminidases (GH101) as well as many uncharacterized sulfatases. These results suggest that these organisms may have the capacity for degradation of complex polysaccharides and glycoproteins to obtain carbon, amino acids, and sulfur, as has been previously suggested for members of this phylum (1). Comprehensive reconstruction of the metabolic pathways encoded in the SAGs will further deepen our understanding of the ecology of these unique *Kiritimatiellaeota* strains in the deep continental subsurface.

**Data availability.** Raw sequencing reads and genome assemblies for the four SAGs have been deposited in the EMBL ENA under project no. PRJEB30981. SAG-specific accession numbers are listed in Table 1. Annotations have been deposited in the Joint Genome Institute's Integrated Microbial Genomes and Microbiomes database (JGI IMG/M) under the accession numbers listed in Table 1.

## ACKNOWLEDGMENTS

We thank the staff of the Bigelow Laboratory Single Cell Genomics Center for the generation of single-cell genomic data. Samples were obtained under scientific research permit DEVA-2013-SCI-0069, issued to D.P.M. from the U.S. National Park Service (NPS), and we thank Richard Friese, Josh Hoines, Genne Nelson, and Kevin Wilson of the NPS and Alisa Lembke of Inyo County, CA, for site access. Thanks also go to Scott Hamilton-Brehm, John Healey, and Brad Lyles of the Desert Research Institute and Jan Amend, Sean Mullin, Victoria Orphan, and Greg Wanger of the NASA Astrobiology Institute's Life Underground project for logistical support.

This work was supported by NASA Astrobiology Institute Cooperative Agreement NNA13AA92A and the U.S. National Science Foundation grants DEB-1441717 and BLOS 19-002. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy contract no. DE-AC02-05CH11231.

## REFERENCES

1. Spring S, Bunk B, Sproer C, Schumann P, Rohde M, Tindall BJ, Klenk H-P. 2016. Characterization of the first cultured representative of *Verrucomicrobia* subdivision 5 indicates the proposal of a novel phylum. ISME J 10:2801–2816. https://doi.org/10.1038/ismej.2016.84.
2. Hugenholtz P, Goebel BM, Pace NR. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol 180:4765–4774.
3. Shepherd ML, Swecker WS, Jr, Jensen RV, Ponder MA. 2012. Characterization of the fecal bacteria communities of forage-fed horses by pyrosequencing of 16S rRNA V4 gene amplicons. FEMS Microbiol Lett 326:62–68. https://doi.org/10.1111/j.1574-6968.2011.02434.x.
4. He Z, Piceno Y, Deng Y, Xu M, Lu Z, Desantis T, Andersen G, Hobbie SE, Reich PB, Zhou J. 2012. The phylogenetic composition and structure of soil microbial communities shifts in response to elevated carbon dioxide. ISME J 6:259–272. https://doi.org/10.1038/ismej.2011.99.
5. Cardman Z, Arnosti C, Durbin A, Ziervogel K, Cox C, Steen AD, Teske A. 2014. *Verrucomicrobia* are candidates for polysaccharide-degrading bacterioplankton in an Arctic fjord of Svalbard. Appl Environ Microbiol 80:3749–3756. https://doi.org/10.1128/AEM.00899-14.
6. Spring S, Brinkmann N, Murrja M, Spröer C, Reitner J, Klenk H-P. 2015. High diversity of culturable prokaryotes in a lithifying hypersaline microbial mat. Geomicrobiol J 32:332–346. https://doi.org/10.1080/01490451.2014.913095.
7. Stepanauskas R, Fergusson EA, Brown J, Poulton NJ, Tupper B, Labonté JM, Becraft ED, Brown JM, Pachiadaki MG, Povilaitis T, Thompson BP, Mascena CJ, Bellows WK, Lubys A. 2017. Improved genome recovery and integrated cell-size analyses of individual uncultured microbial cells and viral particles. Nat Commun 8:84. https://doi.org/10.1038/s41467-017-00128-z.
8. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
9. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
10. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.
11. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu W-T, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Kyrpides NC, Schriml L, Garrity GM, Hugenholtz P, Sutton G, Yilmaz P, Meyer F, Glöckner FO, Gilbert JA, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, The Genome Standards Consortium, Lapidus A, Meyer F, Yilmaz P, Parks DH, Eren AM, Schriml L, Banfield JF, Hugenholtz P, Woyke T. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35:725–731. https://doi.org/10.1038/nbt.3893.
12. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75.
13. Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. J Mol Biol 428:726–731. https://doi.org/10.1016/j.jmb.2015.11.006.
14. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031.
15. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol 57:81–91. https://doi.org/10.1099/ijs.0.64483-0.